2007 AGU Joint Assembly (SM23A-06)

# Magnetospheric Data Discovery and Access

Jan Merka[1,2], T. A. King[3], T. W. Narock[1,2], R. J. Walker[3], S. P. Joy[3]

[1]*University of Maryland, Baltimore County, Baltimore, MD*
[2]*NASA Goddard Space Flight Center, Code 672, Greenbelt, MD*
[3]*Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA*

**http://vmo.nasa.gov**

## VMO Goals
### Provide search capability down to the parameter-value level

0. The two VMO teams (GSFC [PI: Jan Merka] and UCLA [PI: Ray Walker]) closely collaborate and present a ***unified interface*** to users and data providers.

1. The VMO is a distributed data environment with web-accessible graphical user interface (GUI) and application programming interface (API) that provide unified data discovery and retrieval.

2. VMO provides discipline-centric interface and queries for magnetospheric data sets.

3. Flexible implementation: Individual elements of the data environment (data providers, data products, services, etc.) can be introduced as the need arises.

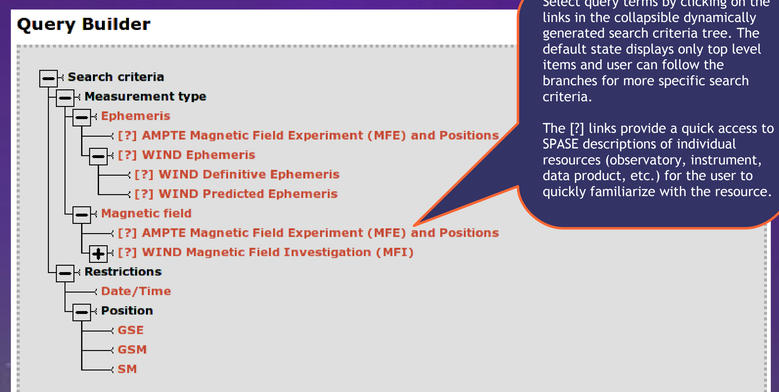
4. Both groups (VMO/G and VMO/U) design and implement VMO Middleware, the *search engine(s)* for magnetospheric data.

5. Develop standardized data product descriptions in collaboration with the SPASE group, other VxOs and the community.

6. Develop and standardize inter-VxO metadata and query exchange mechanism in collaboration with other VxOs. If possible use the same mechanism for VxO-service and/or VxO-data provider communication.

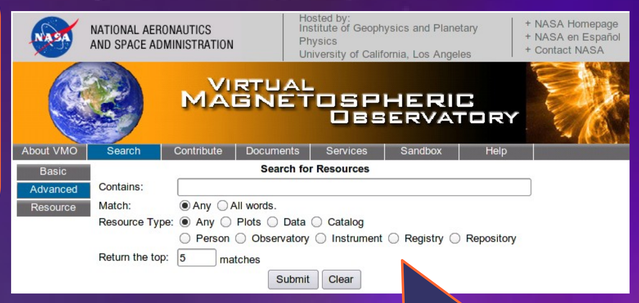
7. Initiate development of add-on VxO services, provide assistance to service providers.

## Prototype Web Search Interfaces and Results Presentation

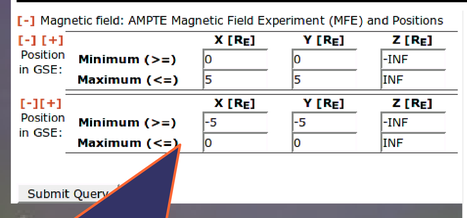



Select query terms by clicking on the links in the collapsible dynamically generated search criteria tree. The default state displays only top level items and user can follow the branches for more specific search criteria.

The [?] links provide a quick access to SPASE descriptions of individual resources (observatory, instrument, data product, etc.) for the user to quickly familiarize with the resource.

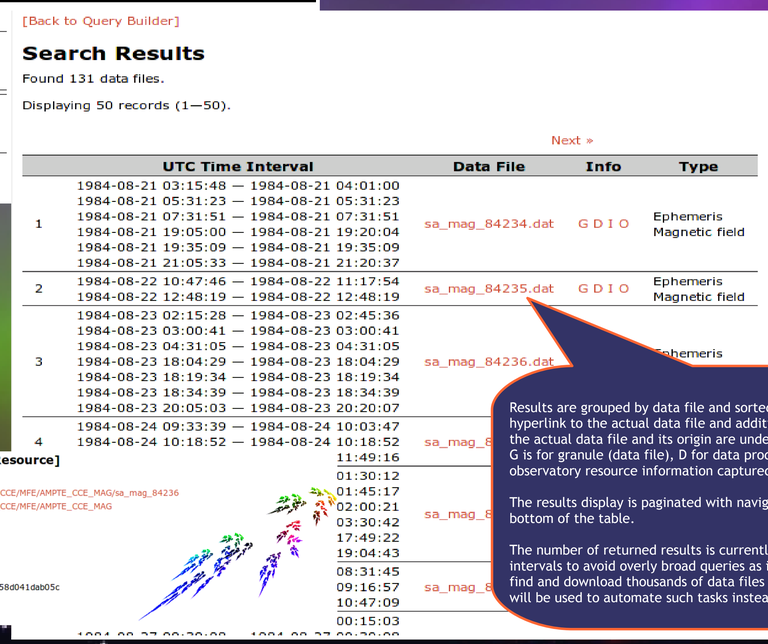



Alternative search interface developed at UCLA which organizes results based on relevance scoring and facets.




Selected search criteria display below the thick horizontal line. Each query term begins with a [-] link that removes the particular item from query. The [+] link provide a quick addition of the same query term, for example to search for multiple time intervals.

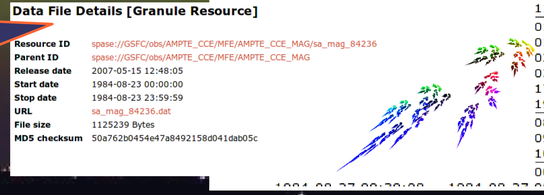
Two submit buttons, at the top and the bottom of the selected query terms, reduce the need for page scrolling.




Results are grouped by data file and sorted by time. The file name is a hyperlink to the actual data file and additional links to information about the actual data file and its origin are under the "Info" column where G is for granule (data file), D for data product, I for instrument, and O for observatory resource information captured by SPASE descriptions.

The results display is paginated with navigation links at the top and the bottom of the table.

The number of returned results is currently limited to 2000 matching time intervals to avoid overly broad queries as it is not expected for users to find and download thousands of data files via the web interface. The API will be used to automate such tasks instead.




Example of data file file information available in SPASE metadata. This can be displayed by following the G hyperlink for each data file.

## Tools and Services

1) SPASE XML Validator

Web-accessible tool that determines XML compliance with SPASE data model.

http://www.spase-group.org/tools/validate/

2) SPASE Registry Server

The SPASE Registry Server is a java application that can harvest resource descriptions expressed in SPASE XML and provide a search service for these descriptions. It can also chain to other registry servers and aggregate all results, returning all matches in a self-organized network of registry servers. It can be run as either a servlet or bean. It uses the SPASE XML Parser package to harvest resource descriptions.

http://www.spase-group.org/tools/registry/

3) SPASE XML Editor

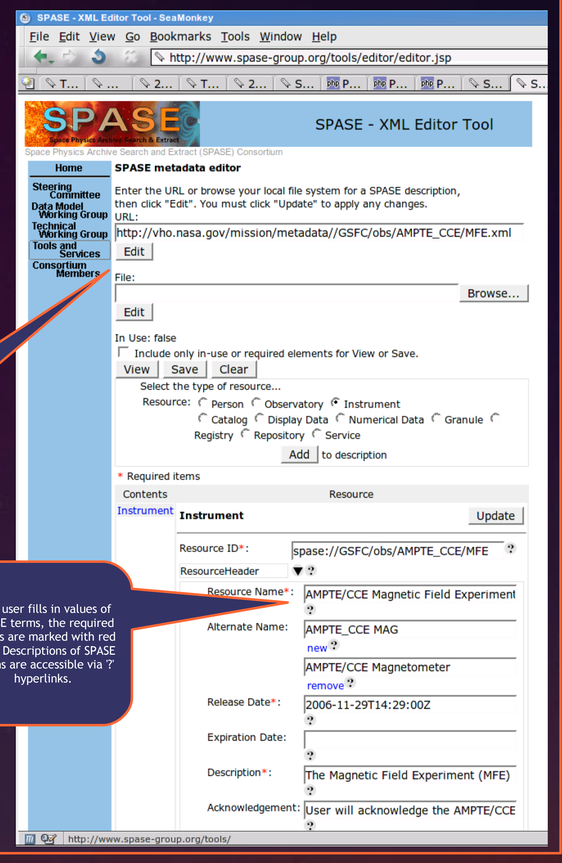
The SPASE XML editor is a web based editor for generating SPASE descriptions. Existing descriptions can be loaded from either a URL or selected from a local file system and uploaded into the editor. Features include viewing and saving of the XML representation of the description.

http://www.spase-group.org/tools/editor/

## SPASE Metadata

The VMO teams are actively involved in SPASE data model development because SPASE is the *interlingua* necessary for understanding, finding, retrieving and exchange of diverse space physics data sets.

This effort resulted in the recent release of the SPASE data model version 1.2.0, a major release that is used for data served by the VMO.




A web-based SPASE XML Editor has been developed for creation and editing of locally or remotely stored XML descriptions.

The user fills in values of SPASE terms, the required terms are marked with red star. Descriptions of SPASE terms are accessible via '?' hyperlinks.

## Summary

1) The VMO has a unique architecture with two centers of activity (NASA/GSFC and UCLA) working on complementary tasks.
2) Several efforts on the architectural components of the VMO are now converging to an interface prototype accessible to users:
- Together with the SPASE consortium, we have finalized SPASE data model version 1.2.0
- Initiated work on developing a database of spacecraft and ground observatory descriptions and their data sets.
- Created and continue developing tools that assist the VxO teams and the community with product descriptions, validation and harvesting.
- Initiated development of Data Query Language that will be used for query specification and result retrieval among VxOs, services and data providers (API).
- Developed early prototypes of data searches using SPASE metadata (web-based search interfaces).

Mark Urwiller©

Contact: Jan Merka (jan.merka@gsfc.nasa.gov)